

Data quality assessment pitfalls

Introduction

Data quality assessments establish if the data is fit for its purpose. It is the foundation for improving data quality. In this article, we will look at what you should not do if you want to get a better understanding of the quality of your data.

A data quality assessment is a process of evaluating data and measuring it against selected quality criteria such as completeness and validity. It also includes analysing the cause and impact of quality problems and sharing the findings. In order to tackle data quality problems in the right way, you need a robust and sound assessment process.

If your assessment is not good enough, you may think that data is good quality when it is not or vice versa. Making the wrong assessment may lead to the wrong actions being taken and that will have an impact on your organisational outcomes.

Lack of planning

Your data quality assessment will be more reliable if it is carefully planned. It is important to involve people with the right skill set and knowledge to carry out the assessment. Technical specialists and process specialists should work in close co-operation to identify, process, test, and refine business rules that are used to confirm and measure the quality of the data.

Before the actual assessment takes place, make sure that you have a good understanding of the data. It is helpful to review any documentation that exists about the data set to be assessed as this will speed up the assessment process.

What you learn from the assessment will be very valuable. It is important that this is documented and saved for future reference. Plan for your assessment to be reproducible. Keeping a record of the assessment process and its findings will allow you to repeat the process. It will become an invaluable source of information for future data quality improvements. This will also maintain your team's knowledge, provide continuity, and minimise risks.

Not considering the purpose of the data

Understanding the purpose can help you identify what to measure, and it is particularly important in communicating quality. If people do not understand the purpose of the data, they will not understand the impact of quality problems. Incorrect data in one part of the data set may not have a material impact on one purpose of the data but it may well throw quality issues for

another use.

A good data quality assessment should be linked to purpose and therefore an assessment must be repeated when a data set is used for another purpose than originally intended.

Measuring the wrong things

Measuring the data that does not matter or that have a low risk not only wastes time and precious resources but also distracts from focusing on fitness for purpose. You do not always have to cover an entire data set. Instead, examine the parts of the data that are important for fulfilling your purpose or purposes.

Focusing on data fields that can be measured rather than what should be measured will result in conclusions about the data that do not address its fitness for purpose. An assessment should focus on what data items you need to be right. This can happen by making sure that you have the people with the right skill set and experience to make that judgement call.

Aiming to be perfect

Getting perfect data is not a realistic goal. Achieving 100% all the time cannot be a requirement. A data set with a single error in it may be good enough to use. There may be more efficiency in coping with data that has been shown to have a few errors than spend time trying to correct every error. What you should be looking for is fitness for purpose and not perfection. Use realistic thresholds for your chosen quality criteria.

Examining data in isolation

Understanding the data lifecycle is important for a meaningful assessment. Data may be subject to errors as they come in. Data can become corrupted during processing. Sometimes data sets are brought together in the wrong way and incorrect assumptions are made when the data is combined. Before you start assessing your data, it is important that you get an understanding of the processes it has undergone to avoid any misleading measurements.

Not being proactive

If you wait passively for a quality issue to be reported before you investigate, it will have already impacted negatively on your organisation. It is much safer to check the quality of your data on a regular basis. This will stop bad quality data from moving through the data lifecycle and allow you to identify issues before they cause greater damage.

The data quality assessment exercise should be a recurring process. The learnings from one assessment will help to review the targets for the next exercise.

You can learn more about data quality assessment from our forthcoming

training course on “Data Quality Action Plans”. The Government Data Quality Hub (DQHub) is developing tools, guidance, and training to help you with your data quality initiatives. Please visit [the Data Quality Hub website](#) for articles, tools, and case studies.

We also offer tailored advice and support across government. Contact us by emailing DQHub@ons.gov.uk.